

## Mental models in propositional reasoning and working memory's central executive

Juan A. García-Madruga, Francisco Gutiérrez,  
Nuria Carriedo, José M. Luzón, and José O. Vila  
*UNED, Madrid, Spain*

We examine the role of working memory's central executive in the mental model explanation of propositional reasoning by using two working memory measures: the classical "reading span" test by Daneman and Carpenter (1980) and a new measure. This new "reasoning span" measure requires individuals to solve very simple anaphora problems, and store and remember the word solution in a growing series of inferential problems. We present one experiment in which we check the involvement of the central executive in conditional and disjunctive inference tasks and compare predictions of the new reasoning span test with those of the classical reading span test. The results of the experiment confirm that reasoning responses, which according to mental model theory require high cognitive work, are predicted by working memory measures. Results also show that some reasoning responses are probably obtained by means of superficial biases or strategies that do not load working memory. The reasoning span test, which involves the central executive to a greater degree, predicts reasoning performance better than the reading span test. The significance and possibilities of the new measure in studying reasoning are discussed.

---

Correspondence should be addressed to Professor Juan A. García-Madruga, Department of Developmental and Educational Psychology, UNED – 28040, Madrid, Spain.  
E-mail: [jmadruga@psi.uned.es](mailto:jmadruga@psi.uned.es)

We thank Phil Johnson-Laird, Ruth Byrne, and Sergio Moreno Ríos for their helpful comments and ideas. The research presented in this paper was begun during the first author's stay in the Department of Psychology at Trinity College, The University of Dublin, visiting Professor Byrne. This visit was funded by the Spanish Ministry of Education and Science; likewise, the present research is part of two projects funded by the Spanish Ministry of Education and Science (PB98-0020-C03-01 and BSO2003-02103).

There are different theoretical views on the cognitive processes that underlie propositional reasoning. One view is that reasoners apply some kind of rules in order to obtain a conclusion; these may be abstract rules of inference that operate in virtue of their form (e.g., Braine & O'Brien, 1998; Rips, 1994), or domain-specific rules of inference (e.g., Fiddick, Cosmides, & Tooby, 2000; Holyoak & Cheng, 1995). A second view is that reasoners rely on imagining possibilities or models (Johnson-Laird & Byrne, 1991, 2002). In spite of their differences, the main theories of propositional reasoning agree that working memory (WM) plays a crucial role in the explanation of people's performance (see, for instance, Johnson-Laird & Byrne, 1991; Rips, 1994).

In this paper we examine the role played by WM's central executive in propositional reasoning by using a new WM test that is based on solving simple pronominal anaphora. We compare a new "reasoning span" test's ability to predict disjunctive and conditional inferences with that of the classical reading span test (RST) of Daneman and Carpenter (1980). The work presented here has been carried out from the perspective of mental model theory and is aimed at clarifying the crucial role it grants working memory in the explanation of propositional reasoning.

## PROPOSITIONAL REASONING: THE MENTAL MODEL THEORY

Mental model theory holds that human reasoning is based on the semantic processes of construction and manipulation of mental models representing the meaning of the premises (e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991, 2002). A main assumption of mental model theory is that working memory plays a crucial role in human reasoning (Johnson-Laird & Byrne, 1991; see also García-Madruga, Gutiérrez, Carriero, Luzón, & Vila, 2005). Given that mental models are constructed and manipulated within WM, reasoners explicitly represent as little information as possible in order to avoid overloading working memory. The construction of mental models is thus guided by the principle of truth: mental models only represent what is true according to the premises, but not what is false (Johnson-Laird, 2000; Johnson-Laird & Byrne, 2002).

The theory allows us to predict and to explain a considerable amount of empirical evidence regarding propositional reasoning. For example, consider reasoning that hinges on sentential connectives, such as "and", "or", and "if". Table 1 summarises the mental models and the fully explicit models of the main sentential connectives.

A conjunction of the form "A and B" means that there is only one possibility in which both "A" and "B" hold, and so the conjunction yields only one model. Hence, inferences based on this connective will be very easy.

TABLE 1  
The initial and fully explicit possibilities or models of some sentential connectives

| <i>Connective</i>       | <i>Initial possibilities<br/>or models</i> |       | <i>Fully explicit<br/>models</i> |       |
|-------------------------|--|-------|----------------------------------|-------|
| A and B                 | A  | B     | A                                | B     |
| A or B                  | A  |       | A                                | not-B |
|                         |  | B     | not-A                            | B     |
|                         | ...  |       | A                                | B     |
| If A then B             | A  | B     | A                                | B     |
|                         |  | ...   | not-A                            | B     |
|                         |  |       | not-A                            | not-B |
| If and only if A then B | A  | B     | A                                | B     |
|                         |  | ...   | not-A                            | not-B |
| Not A unless B          | not-B                                      | not-A | <u>One-way</u>                   |       |
|                         | B  | A     | not-B                            | not-A |
|                         |  | ...   | B                                | A     |
|                         |  |       | B                                | not-A |

“...” denotes a wholly implicit model. Each row represents a model.

Disjunctions of the form “A or B” call for the construction of at least two models, and so inferences based on them should be harder than those based on conjunctions (see García-Madruga, Moreno, Carriero, Gutiérrez & Johnson-Laird, 2001).

Conditional statements such as “If A then B” are compatible with the construction of three possibilities. However, according to model theory, reasoners initially construct only two models like the following:

A      B  
...  
...

The first model accounts for the explicitly expressed possibility that the antecedent (A) and the consequent (B) are both true. The second model is denoted by the three dots and is remembered as a “mental footnote”. It is a fully implicit model that indicates the possibilities in which the antecedent of the conditional proposition is false.

If the mental footnote of the implicit model remains in working memory long enough, people will be capable of fleshing out the implicit model and constructing a fully explicit representation of the previous conditional proposition, as follows:

A      B  
not-A    B  
not-A    not-B

The mental model theory also makes precise predictions for the four rules of inference traditionally posed for conditional assertions (Johnson-Laird & Byrne, 1991; see also Evans, 1993). We consider each in turn.

*Modus Ponens* (MP): from the conditional statement “if A then B” and the categorical premise “there is an A”, the conclusion “there is a B” may be drawn. The situation represented in the categorical premise corresponds to the first model in the initial representation. Therefore, mental model theory predicts that this inference will be easy because the conclusion can be directly extracted from this explicit initial model.

A second valid conditional inference is *Modus Tollens* (MT: “there is no B”, therefore, “there is no A”). Now the categorical premise eliminates the explicit model in the initial representation, demanding a full fleshing out of models. Drawing the not-A correct conclusion requires reasoners to build up a complete representation, in which the categorical premise corresponds to the situation described in the last model *not-A not-B*. MT inferences will obviously be harder because they require the extra cognitive work of fleshing out the models.

The two other inference rules are invalid for the one-way conditionals, though not for two-way conditionals or biconditionals (for instance, “if and only if A then B”). In the affirmation of the consequent (AC: “there is a B”, therefore, “there is an A”), the conclusion may be drawn from the first explicit model. The denial of antecedent (DA: “there is no A”, therefore, “there is no B”) requires fleshing out the models, as with MT inferences. Hence, mental model theory predicts that people will draw more AC than DA inferences.

The study of human deductive ability from conditional assertions has mainly focused on standard factual “if then” expressions. However, the logical meaning of an “if A then B” assertion with neutral content may also be expressed in other ways. For instance, we may use a contrapositive expression such as “if not-B then not-A”, which is logically equivalent to “if A then B”. Likewise, we may use “not A unless B” assertions, which are logically equivalent to “if A then B” and “if not-B then not-A” forms. In this paper, besides disjunctions, we will study two conditional formulations: “if then” and “unless”.

According to philosophers (e.g., Quine, 1972; Reichenbach, 1947) the expression “not-A unless B” is semantically equivalent to “If not-B then not-A”. Nevertheless, some authors have pointed out some differences between these two negative conditionals. For instance, Geis (1973) proposed that the expressions using *unless* could be substituted with *except if*. Likewise, Fillenbaum (1986) claimed that the *only if* connective is the one that best expresses *unless* sentences. Following this idea, from mental model theory, some authors (García-Madruga, Gutiérrez, Carriero, Moreno, &

Johnson-Laird, 2002) have proposed that a “not-A unless B” expression would be initially represented as follows:

|       |       |
|-------|-------|
| not-B | not-A |
| B     | A     |
| ...   |       |

and that the fully explicit representation would also be like either one-way conditionals or biconditionals (see Table 1). Two features of “unless” assertions included in previous representations distinguish them from the standard “if then” expressions: the two initial explicit models, and the backward “B to A” direction of the models. These two characteristics allow us to explain most of the existing evidence that concerns reasoning from “unless” conditional expressions. There is, however, a third main feature of reasoning from “unless” expressions: Previous studies of “unless” show that some individuals make an odd “asymmetric” response. Given, say, premises of the DA form “not-A unless B”, “not-A”, they respond “B”. Similarly, given premises of the AC form “not-A unless B”, “B”, they respond “not-A”. We assume that a superficial “matching” strategy yields these responses (see Carriero, García-Madruga, Gutiérrez, & Moreno, 1999; García-Madruga et al., 2002; Schaeken, García-Madruga, & D’Ydewalle, 1997).

## WORKING MEMORY: THE CENTRAL EXECUTIVE

The concept of WM (Baddeley & Hitch, 1974; see also Baddeley, 1986) has its roots in Atkinson and Shiffrin’s (1968) idea that short-term memory has an active role in cognition: Working memory carries out not only the storage of information but also the simultaneous processing of information. Baddeley and Hitch’s structural model of WM consists of three components: (1) a central executive that has a limited capacity, and controls and coordinates the two other “slave” components, which are (2) the phonological loop, and (3) the visuo-spatial sketchpad. This classical version has recently been modified by including a fourth component—the *episodic buffer*—whose main function is to recover and integrate information from long-term memory (see Baddeley, 2000).

In Baddeley and Hitch’s (1974) initial model of WM, the role of the central executive was quite vague and imprecise, and focused principally on the coordination of the two slave systems. Later, Baddeley (1986, 1990) incorporated some ideas from Norman and Shallice’s (1980; see also Shallice & Burgess, 1993) theory of a supervisory activating system into his conception of a central executive. Thus, by enhancing the role of attentional control, he postulated three primary functions of the central executive: focusing attention, switching attention, and activating representations

within LTM. An interesting corollary of the supervisory function of the central executive is its capacity to inhibit automatic processes and to discard irrelevant information (see Baddeley, 1996; Baddeley & Logie, 1999; Engle, Kane, & Tuholsky, 1999; see also Handley, Capon, Beveridge, Dennis, & Evans, 2004). In this work we propose that in order to get good measures of the central executive we have to use tasks that imply the need to inhibit automatic processes and discard irrelevant information.

During the two last decades, there has been a pattern of increasing attention paid by researchers to the central executive and its crucial role in cognitive functioning (see, for instance, Richardson, 1996). Related to the increasing role conferred on the central executive, some researchers have stressed the two mutually related components of working memory: storage capacity and processing effectiveness (see Just & Carpenter, 1992). One of the main contributions of this perspective is the study of individual differences that use measures such as the reading span test (RST) developed by Daneman and Carpenter (1980). Since the RST requires participants to read aloud a series of sentences and to recall the final word of each sentence, this test is able to assess both storage capacity and processing effectiveness, the two important components of working memory. The processing part of the task is reading aloud; the storage part of the task is recalling the final words. Since the tasks require some kind of attentional control, RST is a measure of the central executive in working memory (see Engle & Oransky, 1999; Whitney, Arnett, Driver & Budd, 2001). This is the main interpretation of the high correlations found between RST and different reading comprehension and verbal reasoning abilities (Daneman & Carpenter, 1980; Just & Carpenter, 1992; Whitney, Ritchie, & Clark, 1991). There is, however, a possible criticism of this interpretation based on a circularity argument: there is some overlap between reading processes involved in the RST and verbal ability measures.

## MENTAL MODEL THEORY AND EXECUTIVE CONTROL

The mental model theory has motivated researchers to study the different components of WM and their relative involvement in reasoning. It is sometimes supposed that mental models are a kind of “iconic” or image-based representation, and therefore that the visual component of WM should be more involved than the verbal one. Accordingly, Johnson-Laird (1985, p. 190) stated that “the ability to construct alternative models ... should correlate with spatial ability rather than verbal ability”. However, more recently Johnson-Laird shows that mental models are types of semantic representations distinct of propositional representations and images: “Mental models are iconic as far as possible but certain components of them are necessarily symbolic” (Johnson-Laird, 2004, p. 203). Therefore,

neither the visual component of working memory nor the verbal one is crucial for reasoning. For the crucial component, what we need to look at is the central executive.

According to mental model theory, propositional reasoning is a complex cognitive activity that requires the processing, activation, and maintenance of relevant information, and the moment-to-moment control and monitoring of the process of finding a solution. Therefore, the three main functions of CE—the capacity to focus attention, to switch attention, and to activate representations within LTM—are clearly required in order to draw a conclusion in a propositional inference problem, particularly if the problem requires the construction and manipulation of multiple models. During the process of solving deductive inferences, reasoners have to (1) focus their attention on diverse cognitive tasks, interpret each premise, integrate the meaning of both premises, and search for possible counterexamples in order to formulate conclusions (see, for instance, Johnson-Laird & Byrne, 1991); (2) shift their attentional focus from one cognitive task to another; and, in order to construct semantic representations from the meaning of the premises, they also have to (3) activate and use their knowledge stored in LTM. Therefore, there should exist a close relationship between CE and propositional reasoning performance.

There is some evidence in favour of the hypothesis that CE is involved in reasoning. A number of studies in WM and reasoning have been carried out over the last two decades in order to examine the involvement of WM in diverse reasoning tasks, mainly syllogistic and propositional reasoning tasks (see Gilhooly, 1998). These studies can be divided into two main groups: those using the dual-task procedure and those based on a correlational approach.

In Baddeley and Hitch's (1974) double-task procedure, whereas the primary task is a reasoning task, the secondary task loads one of the WM components—the central executive, the phonological loop, or the visuo-spatial sketchpad. The rationale is simple: the more the secondary task affects reasoning, the greater the involvement of the WM component in question. Using the dual-task procedure, Toms, Warris, and Ward (1993) also found clear evidence for the role of the central executive in conditional inferences, but no evidence for the involvement of the visual sketchpad or the phonological loop. In other words, the central executive load resulted in poorer conditional reasoning, particularly on modus tollens inferences, the most difficult sort.

Likewise, Klauer, Stegmaier, and Meiser (1997) compared how different secondary tasks affect the three WM components for both propositional (conditionals and disjunctions) reasoning and spatial reasoning. These authors found that, although spatial reasoning was clearly more disrupted by the secondary tasks affecting the central executive, propositional reasoning

was also affected. Klauer and his colleagues interpreted their results according to Evans' two-stage theory of reasoning (Evans, 1984, 1989; Evans & Over, 1996). In this theory, some initial preconscious or "heuristic" processes select relevant information, and thereby direct and focus the workings of a subsequent explicit, "analytic" stage. According to Klauer and his colleagues, in propositional reasoning participants use a heuristic strategy that is based on linguistic surface characteristics and that consumes hardly any cognitive resources. For this reason, the initial heuristic processing assumes a smaller load on working memory for propositional reasoning than for spatial reasoning.

As Klauer et al. (1997) hold, their results may be understood better in light of work on working memory and syllogistic reasoning by Gilhooly and collaborators (Gilhooly, Logie, Wetherich, & Wynn, 1993; Gilhooly, Logie & Wynn, 1999). These authors have found a clear relation between working memory load and the kinds of strategies used by participants in solving categorical syllogisms. The accuracy of high-skilled participants was affected by secondary tasks loading WM; the performance of low-skilled participants was not affected, probably because instead of reasoning they used surface strategies that demanded fewer cognitive resources.

As we have just seen, two main points may be drawn from the studies carried out within the dual-task procedure: (1) the crucial role of the central executive on propositional reasoning tasks, and (2) the probable existence of surface strategies or heuristics that keep reasoners from overloading their working memory.

From the correlational perspective, on the other hand, recent studies have obtained only weak and unreliable correlations between WM capacity and conditional reasoning responses. Among these studies, Markovits, Doyon, and Simoneau (2002) gave reasoners the four conditional inference problems with concrete and abstract content. As measures of WM they used a verbal WM capacity test based on that of Daneman and Carpenter (1980) and a visuo-spatial WM task. They found a number of significant correlations between reasoners' performance for two kinds of content and WM tests, but the correlations were rather low (around .20). Moreover, the highest correlation found by Markovits and colleagues was between verbal WM and "modus ponens" (MP) inferences (0.32), while no significant or even positive correlations with "modus tollens" (MT) inferences were obtained. These findings are particularly unexpected, since MT inferences require more cognitive work, and are consequently more difficult than the direct MP inferences.

The results of work by Barrouillet and Lecas (1999) are more optimistic. In their study, participants were three groups of children and young adolescents who were asked to list all the cases compatible with a conditional rule. Working memory capacity was assessed using Case's

(1985) counting span task. The authors found a reliable and high positive correlation between WM capacity and conditional reasoning (0.648). However, it may be impossible to generalise these results to other tasks or to populations other than children. As diverse studies have shown, the correlation between WM capacity and language comprehension tends to be higher for children than for adults (see García-Madruga, Gárate, Elosúa, Luque, & Gutiérrez, 1997; see also Gathercole & Baddeley, 1993).

As in studies carried out following the dual-task procedure, studies with syllogistic and spatial reasoning problems have borne out the crucial role of the central executive (see also Capon, Handley, & Denis, 2003). However, in propositional reasoning, the results of the correlational studies, such as that of Markovits et al. (2002), are neither clear nor conclusive. This might be due to the kind of CE measures used. In our opinion, most of the measures, in particular the RST, may not load the CE enough. RST is certainly a measure of CE: It is a dual-task measure that requires shifting the focus of attention from one task (the processing component: reading aloud some sentences) to another (the storage component: remembering the last word of each sentence). However, the first task is automatic, since the word that has to be stored and remembered is arbitrarily chosen. Our aim in this paper is to present and check a new measure that increases the involvement of CE by means of increasing the demands of attentional control.

Our new measure is based on the solving of anaphora problems. This new “Reasoning Span test” introduces some modifications to the processing component, while maintaining constant the storage component: participants have to read a sequence of sentences and then *infer* a word that must be remembered. Unlike RST, in our new measure participants have to inhibit and discard some alternative word solutions and select the correct word that they have to remember. We have also designed and checked another Reasoning Span test based on the solving of very easy analogy problems. In the Analogy test, participants have to read a sequence of verbal analogy problems and then infer the word solution of each problem that must then be remembered (see García-Madruga et al., 2005).

Just like the aforementioned use of RST in predicting text comprehension, there is also some kind of circularity in using our new “Reasoning Span Tests” for predicting reasoning performance (see Philips & Forshaw, 1998). For this reason we selected a very simple and daily inferential task for our measure of CE that is quite different from deductive reasoning tasks: pronominal anaphora. In order to check the minimal difficulty of anaphora problems, they were tested with a sample of 35 university participants. The percentage of erroneous responses to anaphora problems was 1.84%. This means that the inferential task of solving the anaphora problems does not allow the prediction of deductive reasoning performance.

## THE EXPERIMENT

The main objective of this experiment is to check the involvement of the central executive in propositional reasoning and to test the predictive capacity of a new measure of CE. The new test clearly increases the demands of attentional control and, therefore, should provide a better measure of the central executive's capacity for reasoning. In this sense, although we expect the new test to be harder than the standard RST, there should nevertheless be correlations among the various working memory measures.

In order to examine the relation between propositional reasoning and the two WM measures, besides the two WM tests, participants also carried out a reasoning test that included three sentences—two conditionals and one disjunction. We chose these propositional sentences for two main reasons: (a) according to model theory, in order to solve most inferences participants need to build multiple models; (b) some responses will likely be yielded by means of superficial bias or strategies.

One conditional was of the form “if A then B” and the other was of the logically equivalent form “not-A unless B” (see García-Madruga et al., 2002). The participants had to generate their own conclusions to the four classical inferences (MP, DA, AC, and MT) for both sorts of conditionals. The disjunction was of the form “A or B, or both”, and the participants had to generate a conclusion given the additional four categorical premises: A, B, not-A, and not-B, respectively.

According to the mental model theory, reasoners make most inferences from mental models (see Table 1). Therefore, the theory predicts that for “if then” inferences, MP and AC should be more frequent than MT and DA. On the other hand, as we said above, people reasoning from “unless” expressions construct from the outset two explicit models. From this two-model initial representation, reasoners can draw the four inferences “MP”, “MT”, “AC”, and “DA”. For “unless” conditionals, however, there should be some superficial “asymmetric” responses which we referred to in a previous section. These superficial matching responses do not consume cognitive resources and, therefore, we may surmise that these responses will be more usual within people with low WM scores.

Finally, an “unless” conditional assertion such as “You won't pass the exam unless you study harder” tends to be interpreted as a biconditional; in other words, people probably understand this assertion to mean that you will pass the exam “if and only if” you study harder. As a consequence of the biconditional interpretation of “unless”, the four conditional inferences are all valid.

The inclusive disjunction “A or B, or both” requires a more detailed analysis. Its fully explicit models are as follows (see Table 1):

|       |       |
|-------|-------|
| A     | B     |
| A     | not-B |
| not-A | B     |

This representation is highly demanding, and hence individuals are likely to build only the models they need to reach the conclusion from each premise (Morris & Sloutsky, 2002; Richardson & Ormerod, 1997). Thus, given the categorical premise A, reasoners should build the two models:

|   |       |
|---|-------|
| A | B     |
| A | not-B |

from which they can make the correct response: “There is no valid conclusion”. A similar process can be hypothesised for the categorical premise B, for which reasoners should build the models:

|       |   |
|-------|---|
| A     | B |
| not-A | B |

that also yield the correct “no valid conclusion” response. However, with these two affirmative categorical premises (A and B) some people may give an invalid “symmetric” conclusion by applying some kind of superficial matching strategy. In this way, from “A” they answer “B” and from “B” they answer “A”. These superficial responses imply a complete misunderstanding of the meaning of disjunctions. As for “asymmetric” erroneous responses for “unless”, these superficial matching responses do not consume cognitive resources and, therefore, we may surmise that they will be more usual within people with low WM.

For the two negative categorical premises not-A and not-B, the reasoning process may be different. Reasoners probably have to construct the three models in order to reach the valid asymmetric conclusions: B from not-A, and A from not-B. These models should block the symmetric responses not-A, therefore not-B, and not-B, therefore not-A.

In prior work, our research team used an experimental design in which we examined the propositional reasoning ability of High and Low RST working memory participants. The results of this study confirmed the differences between High and Low working memory participants in conditional reasoning (Meilán, García-Madruga, & Vieiro, 2000). In the current experiment we used the same design in order to compare the propositional

reasoning ability of High and Low working memory participants in two WM measures: RST and the Anaphora test.

Our hypotheses were two. First, we expected correlations between WM and Reasoning measures. Positive correlations should occur between the WM measures and those responses deriving from a greater cognitive load. Hence, for “if then” conditionals there should be positive correlations with responses that require fleshing out the models: MT and DA inferential responses, and correct “not valid conclusion” responses for AC and DA. For “unless” conditionals there should be positive correlations with all four inferences that require the construction of two initial models. Granting the usual biconditional interpretation of “unless”, all four inferences are valid. Likewise, we predicted a positive correlation among the WM measures and the four valid responses to the inclusive disjunction. On the other hand, we predicted a negative correlation among WM measures and the responses yielded by superficial matching strategies in “unless” and in disjunctions. The main correlational prediction, however, is that the anaphora measure of working memory should yield higher and more reliable correlations with inferential performance than the standard reading span measure (RST).

Our design also allows us to experimentally check a second hypothesis: High WM groups will give more responses deriving from high cognitive load, and Low WM groups will give more responses yielded by superficial matching strategies. Likewise, our second hypothesis includes the main prediction that the anaphora measure of working memory should predict reasoning performance better than the standard reading span measure (RST).

## METHOD

### Design and participants

A total of 104 students from a psychology class at the “UNED” of Madrid and Calatayud (Zaragoza) participated voluntarily in the experiment for course credit. The participants had received no training in logic and had not been previously tested in any experiment on working memory and reasoning. A complete within-subjects design was used: All participants carried out the two working memory tests and the reasoning test. A total of 12 participants were eliminated for not having adequately answered at least one of the WM tests. From the whole group of participants we selected from each of the WM tests two groups consisting of High and Low WM participants: High WM participants scored at least a standard deviation above the mean, Low WM participants scored at least a standard deviation below the mean.

## Tasks and materials

### *Working memory*

*RST.* The Reading span task (Daneman & Carpenter, 1980; adapted Spanish version; see Elosúa, Gutiérrez, García-Madruga, Luque, & Gárate, 1996) was used. Participants carried out the task collectively, in groups of 25. Each participant had to read to himself a series of progressively increasing phrases presented on a screen, and was then asked to recall the last word of each phrase and write each of them down in the correct order. The test contained 60 unrelated sentences. Each participant began by reading two sentences, and then three sentences, and continued reading progressively four, five, and six sentences. This series was repeated three times, until each participant had read all 60. Each sentence appeared on a single line in the centre of the screen and was 12 to 14 words in length, the last word containing two or three syllables.

After doing this, half of the series randomly included a question in which participants were asked whether a word had or had not been presented in previous phrases. In this way, participants were led to read the phrases in their entirety and not only the last word of each.

*Anaphora test.* We used a new version of the Anaphora test designed by our research team (for a complete presentation of the anaphora problems, see Gutiérrez, García-Madruga, Carriero, Luzón, & Vila, 2005). People have to remember a word that is the referent of a simple pronominal anaphora. There were two possible words from which to choose. Here is an example:

- Robert painted **it** white before the summer arrived.  
 – roof  
**– façade**

The word in bold, **façade**, is the correct response because it matches the meaning of the sentence and the **gender** (feminine) of the pronoun (in Spanish: *Roberto la pintó de blanco antes de que llegara el verano. –tejado – fachada*). The foil is semantically appropriate but grammatically inappropriate. As we can see in the above example, our anaphora problems present the pronoun before the noun: they are hence backward anaphora or “cataphora” problems (see, for instance, Garnham, 2001).

The task was also carried out collectively, in groups of 25 participants. They had to read to themselves a series of progressively increasing anaphora problems presented on a screen, and then recall the word solution of each anaphora problem and write each of them down in the correct order. As in

RST, half of the series randomly included a question in which participants were asked whether a word had or had not been presented in previous anaphora problems.

The new tasks consist of 42 inference problems through three series of different levels of two, three, four, and five problems each. After reading the instructions, participants trained by solving three trials of two problems each, and then began the experiment by doing a different three trials of two problems each, followed by three trials of four problems each, until three trials of five problems each had been reached. Level 6 was removed because a pilot study showed that no participant was able to reach a consistent performance in level 5. The order of appearance and familiarity of the word solutions was also controlled.

The scoring procedure of WM tests was that developed by Elosúa and collaborators (1996) for RST. This procedure scores the number of words that participants are able to remember with minimum consistent performance. In each of the three series in each level, participant performance can be (1) correct (accurate words, correct order), (2) half correct (accurate words, incorrect order), and (3) incorrect. The minimum consistent performance in each level is reached when a participant gets at least half of the maximum performance in that level:

- (a) Three series of words half correct.
- (b) One series of words correct, one half correct, and one incorrect.

Every performance better than the minimum consistent performance in the same or higher levels was scored by the addition of decimals. In the same level, each supplementary correct response would add two decimal points and each supplementary half correct response one decimal point. For instance, minimum correct performance in the fourth level is 4, and maximum performance in the fourth level is 4.3; if a participant remembers accurately and in the correct order only two of the three series of four words (level 4), his scoring will be 4.1. In a higher level, a supplementary correct response would add 5 decimal points and a supplementary half correct response 4 decimal points. For example, if the previous participant also remembers a series of five words (level 5), but in the incorrect order, his scoring will be  $4.1 + .4 = 4.5$ . (see Gutiérrez et al., 2005).

### *Reasoning task*

We constructed three sets of problems, one based on “if then” conditionals, one based on “unless” conditionals, and one based on “A or B, or both” disjunctions. Each set contained 12 problems: three items in each of four inferences. We used neutral content with information about persons and

TABLE 2  
The three assertions and the four categorical premises used in the experiment

*Assertions*

If Óscar is in Granada then Nuria is in Sevilla.  
Óscar is not in Granada unless Nuria is in Sevilla.  
Óscar is in Granada or Nuria is in Sevilla, or both.

*Categorical premises*

|                         |      |
|-------------------------|------|
| Óscar is in Granada     | (MP) |
| Nuria is in Sevilla     | (AC) |
| Óscar is not in Granada | (DA) |
| Nuria is not in Sevilla | (MT) |

locations (e.g., Óscar is in Granada or Nuria is in Sevilla, or both). Each problem consisted of a propositional premise (either conditional or disjunctive) and a categorical premise corresponding to the four logical possibilities: A (MP), B (AC), not- A (DA), and not-B (MT) (see Table 2). Participants had to generate their own conclusions. The problems were in the participant's native Spanish and referred to common Spanish proper names and well-known Spanish cities.

### Procedure

First, participants carried out the two working memory tests in a silent room in groups of about 25. The pace of presentation of the sentences or problems on the screen, as well as the time granted to the recall task, was previously established from a pilot study. RST was always carried out before the Anaphora test. Afterwards, all the participants carried out the paper and pencil reasoning test.

## RESULTS AND DISCUSSION

Participants' performance in the two WM tests was significantly different. The mean score in the Reading Span Test (4.13 words; 2.1–6.1) was reliably higher than in the Anaphora test (mean 3.67 words; 2.0–5.3; *t*-test;  $p < .0001$ ). The Pearson correlation between the two WM measures was positive and significant: 0.60 ( $p < .001$ ).

As the results show, the introduction of simple inference tasks into a test of working memory significantly increases difficulty. The work of inferring the word that has to be remembered increases WM load and reduces the resources for storing the result.

Table 3 presents the percentages of diverse responses for the three sorts of propositional inferences. Table 4 presents the correlations among the two measures of WM and the correct responses, for the whole group ( $N=92$ ).

TABLE 3  
Percentages of responses for both conditional statements as a function of inference type (correct responses in bold)

| <i>F</i>               | <i>Categorical premise</i> |                        |                            |                            |
|------------------------|----------------------------|------------------------|----------------------------|----------------------------|
|                        | <i>A</i> ( <i>MP</i> )     | <i>B</i> ( <i>AC</i> ) | <i>not-A</i> ( <i>DA</i> ) | <i>not-B</i> ( <i>MT</i> ) |
| <i>If A then B</i>     |                            |                        |                            |                            |
| Symmetric responses    | <b>97</b>                  | 72                     | 45                         | <b>58</b>                  |
| No conclusion          | 1                          | <b>26</b>              | <b>54</b>                  | 40                         |
| Asymmetric responses   | 0                          | 0                      | 0                          | 1                          |
| Other                  | 2                          | 2                      | 1                          | 1                          |
| <i>Not A unless B</i>  |                            |                        |                            |                            |
| Symmetric responses    | <b>88</b>                  | <b>82</b>              | <b>64</b>                  | <b>80</b>                  |
| No conclusion          | 10                         | 6                      | 18                         | 15                         |
| Asymmetric responses   | 1                          | 11                     | 16                         | 5                          |
| Other                  | 1                          | 1                      | 2                          | 0                          |
| <i>A or B, or both</i> |                            |                        |                            |                            |
| Symmetric responses    | 28                         | 25                     | 9                          | 9                          |
| No conclusion          | <b>58</b>                  | <b>63</b>              | 43                         | 45                         |
| Asymmetric responses   | 9                          | 8                      | <b>43</b>                  | <b>44</b>                  |
| Other                  | 5                          | 4                      | 5                          | 2                          |

TABLE 4  
Correlations between WM and diverse reasoning measures (*N*=92)

| <i>Measures of WM</i> | <i>Correct</i> | <i>Correct</i> | <i>Correct</i> | <i>Overall</i> | <i>Multiple</i> | <i>Superficial</i> |
|-----------------------|----------------|----------------|----------------|----------------|-----------------|--------------------|
|                       | <i>If then</i> | <i>Unless</i>  | <i>or</i>      | <i>correct</i> | <i>models</i>   | <i>erroneous</i>   |
| Reading span (RST)    | .22*           | .12            | .10            | .18*           | .16             | -.09               |
| Anaphoric measure     | .26**          | .30**          | .23*           | .36**          | .22*            | -.22*              |

\*\**p* < .01; \**p* < .05.

The table also shows the correlations among the three WM measures and, on the one hand, responses requiring multiple models and, on the other, those erroneous responses that according to our theoretical proposal can be derived from a superficial “matching” strategy. The kinds of responses included in these categories are as follows:

(A) Multiple model responses:

- DA invalid and MT valid inferences, and DA and AC “No valid conclusion” responses for “if then”.
- The four valid inferences for “unless”.
- The four valid responses for “or”.

## (B) Superficial erroneous responses:

- DA and AC asymmetric responses for “unless”.
- Symmetric *A*, therefore *B* and *B*, therefore *A* inferences for “or”.

In general, the predictions were confirmed. The first three columns of Table 5 show a number of positive correlations among WM measures and reasoners' performance on the three sentences: “if”, “unless”, and “or”. The correlations were always higher for the Anaphora measure than for the RST measure. This pattern can be observed clearly in the last three columns, in which the correlations with overall correct responses, multiple model responses, and superficial erroneous responses are shown. The most conspicuous results are probably those in two last columns: The positive correlations between WM measures and multiple model responses and the negative correlations between WM measures and erroneous responses involving a small load on WM; that is, those yielded by a superficial “matching” strategy. Although all the correlations are in the predicted direction, only the correlations with Anaphora are reliable.

In order to have a clearer view of these results, a regression analysis on the three main variables, overall Correct Responses, Multiple Model Responses, and Erroneous Superficial Responses, was carried out. Multiple regression analysis showed that the two measures of WM as an independent variable explains 13% of the variance of Correct Responses,  $F(2, 89) = 6.82$   $p < .002$ , although the only significant variable was WM Anaphora ( $B = .08$ ,  $\beta = .40$ ,  $p < .002$ ). Multiple regression on Multiple Models was

TABLE 5  
Percentages of diverse reasoning responses for low and high WM groups in Reading span and Anaphoric measures, and *t*-values for the differences

|                                | <i>Correct If then</i> | <i>Correct Unless</i> | <i>Correct or</i> | <i>Overall correct</i> | <i>Multiple models</i> | <i>Superficial erroneous responses</i> |
|--------------------------------|------------------------|-----------------------|-------------------|------------------------|------------------------|--|
| RST Low ( <i>N</i> = 19)       | 56                     | 69                    | 47                | 57                     | 47                     | 27                                     |
| RST High ( <i>N</i> = 17)      | 70**                   | 86*                   | 62                | 73**                   | 63*                    | 15                                     |
| <i>T</i> values                | −2.58                  | −1.94                 | −1.36             | −2.71                  | −1.89                  | 1.34                                   |
| Anaphora Low ( <i>N</i> = 21)  | 55                     | 58                    | 46                | 53                     | 48                     | 29                                     |
| Anaphora High ( <i>N</i> = 16) | 68*                    | 86**                  | 68*               | 74**                   | 64*                    | 13*                                    |
| <i>T</i> values                | −2.14                  | −3.97                 | −1.80             | −3.42                  | −1.76                  | 1.85                                   |

Participants in low or high working memory groups scored at least a standard deviation below or above the mean.

\*\* $p < .01$ ; \* $p < .05$ .

not reliable,  $F(2, 89) = 2.23, p < .11$ , although WM Anaphora regression was ( $B = .06, \beta = .22, p < .04$ ), explaining only 4.7% of variance. Multiple regression analysis on Erroneous Superficial Responses was also not significant,  $F(2, 89) = 2.42, p < .1$ , whereas WM Anaphora regression was, explaining almost 5% of variance ( $B = -.06, \beta = -.22, p < .04$ ).

Table 5 presents the results of the diverse reasoning measures from High and Low working memory groups. High WM RST participants ( $N = 17$ ) gave reliably more correct responses in two of the three types of assertions, as well as in overall correct responses and in multiple model responses, than did low WM RST participants ( $N = 19$ ). However they did not give reliably fewer superficial erroneous responses. On the other hand, as expected, High WM Anaphora participants ( $N = 16$ ) gave a reliably different proportion of correct responses than low WM Anaphora participants ( $N = 21$ ) for all the reasoning responses, including a smaller proportion of superficial erroneous responses.

The results illustrate the key role of working memory in explaining propositional reasoning abilities, and particularly the role played by the central executive. The new WM test—the Anaphora test—predicted participants' reasoning performance better than the RST. This was confirmed by the higher and more reliable correlations and the significant regression with the Anaphora test. The most important result, however, refers to the difference between High and Low WM groups and, particularly, that when comparing High and Low WM Anaphora groups all the differences were reliable. Both measures of working memory's CE are able to predict better performance on multiple model problems, but only the Anaphora test is able to predict a smaller proportion of superficial erroneous responses.

## GENERAL DISCUSSION

In this paper we have claimed that CE plays a crucial role in the explanation of participants' reasoning performance. We have also confirmed this proposal. As a measure of the capacity of the central executive we used the classical RST and a new measure. The new measure introduces a trivial inferential problem as part of the processing required in the test. The introduction of this new inferential component is presumably responsible for the improvement—in comparison with the classical reading span task (RST)—in the predictions of reasoning performance: The participants are forced to regulate their cognitive resources in a complex task that requires both inference and memory.

We have proposed the use of differences between individuals in studying working memory and reasoning. The process of reasoning implies the building, integration, and manipulation of semantic representations and hence demands high cognitive resources, particularly in the control and supervisory functions of the central executive. However, there are some responses in “unless” and “or” assertions that require in their explanation a

superficial strategy. These superficial responses are automatic and do not consume cognitive resources (see Gilhooly, 1998).

As we have shown, there are positive correlations between WM and reasoning responses that require high levels of mental work, and negative correlations between WM and reasoning responses that require low levels of mental work. The positive correlations with correct responses, and especially with those that require the representation of multiple models, provide support for the model theory.

Our aim in this paper was also to examine the relation between working memory's CE and propositional reasoning, by using a new test. Our new Reasoning Span Test depends on a secondary task calling for an inferential decision. In place of materials that participants arbitrarily select, such as the last word of a sentence, the items that individuals now have to recall are a result of an inference that forces participants to inhibit and discard alternative word solutions. Thus, two primary results in favour of our hypotheses deserve more attention: the reliably higher difficulty of the Reasoning Span Test than the Reading Span Test (RST), and additionally their higher capacity in predicting inferential reasoning.

As analysed in the Introduction, there might be an element of circularity when using measures of WM to predict reasoning ability when the measure itself calls for some kind of reasoning. There are two main interrelated arguments against the circularity of the Reasoning Span measure. First, as in RST, participants' scores in the Reasoning Span test are real working memory capacity measures, not inferential or language understanding measures: we measured the number of words that participants are able to remember. Second, the pronominal anaphora problems we used in the test are quite different from the propositional reasoning inferences that we have investigated. These pronominal anaphora inferences are normally made automatically on-line, during the process of language comprehension, in order to obtain the local coherence of discourse (see, for instance, Garnham, 2001; Gernsbacher, 1990). Moreover, the backward anaphora inferences in the Reasoning Span test were selected for the very reason that they were easy. In fact, they are trivial: Spanish readers have an unambiguous morphological index—the gender of the pronoun—to solve the anaphoric reference.

Thus, it is not the demands of reasoning in solving the anaphora problems that increases the predictive power of the new measures. Instead, the increased ability of the new measures in predicting reasoning performance comes from the need to shift the focus of attention from one task to the other. The new secondary task (anaphoric inference) requires more attentional control resources and thereby forces people to increase their cognitive work in constructing and manipulating representations. Likewise, in the new reasoning span task, participants have to inhibit and discard erroneous word solutions to anaphora problems.

These two features of the new working memory task, the need to change the focus of attention and to discard or inhibit some direct erroneous responses, are typical of cognitive tasks in which the central executive is involved and are common to propositional reasoning tasks. It is hence unsurprising that our new reasoning span test is harder than RST, and that it is able to predict better a task in which CE is as highly involved as it is in propositional reasoning.

Perhaps the most interesting suggestion arising from the foregoing study is that the use of WM measures can help clarify the diverse cognitive processes through which reasoners reach a conclusion. We use the case of the hypothesised “superficial” responses to illustrate this point. Our results establish that anaphora low working memory individuals tend to commit more superficial errors: the asymmetric responses to “not-A unless B” DA and AC inferences, on the one hand, and the symmetrical responses to the affirmative inferences “p therefore q” and “q therefore p” in “A or B, or both” assertions, on the other. This clearly confirms our hypothesis regarding the use of a superficial matching strategy.

Likewise, the regression results for superficial erroneous responses with anaphora, as well as the negative correlations found between these kinds of responses and the diverse measures of WM (especially with anaphora) seem to confirm that they do not require many cognitive resources. We may hence conclude that these responses are probably reached by means of a superficial matching strategy. Therefore, our results support the idea of an automatic, superficial linguistic representation of premises that underlies one’s ability to reason with linguistic assertions, and that may therefore at times affect how reasoners draw conclusions.

The results found enhance the view that reasoning operates as a dual process, characterised in different ways by diverse authors: Superficial linguistic vs semantic (García Madruga, 1983), heuristic vs analytic (Evans, 1984), tacit vs explicit (Evans & Over, 1996). These accounts maintain that people’s logical abilities are based in the development of Type 2 processes—semantic, analytic, or explicit—and that their capacity to inhibit responses are produced by Type 1 processes—superficial, heuristic, tacit. As we argued in the introduction, the ability to over-ride the Type 1 system and inhibit automatic heuristic-based responses is related to working memory’s central executive (see also Stanovich, 1999). High WM central executive individuals will hence be more able to resist these superficial responses and instead rely on the semantic process of constructing models of the meaning of premises, from which they will be able to draw correct conclusions. The Anaphora test confirms that high WM central executive participants are able to inhibit superficial responses, and these results show again that the Reasoning Span test is able to predict people’s performance better than the RST.

The results of this work also show that a deeper exploration of the new Reasoning Span measures is in order, in particular, how reliable our new measures are in predicting higher cognitive processes such as reasoning and reading comprehension, and what sorts of methodological improvements we can make regarding their capacity to predict such cognitive processes (see Friedman & Miyake, 2004).

Manuscript received 6 February 2006

Revised manuscript received 20 September 2006

First published online 18 April 2007

## REFERENCES

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.
- Baddeley, A.D. (1986). *Working memory*. Oxford, UK: Oxford University Press.
- Baddeley, A. D. (1990). The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 54–73). Cambridge, UK: Cambridge University Press.
- Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology*, 49A, 5–28.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
- Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge, UK: Cambridge University Press.
- Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking and Reasoning*, 5(4), 289–302.
- Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking & Reasoning*, 9, 203–221.
- Carriero, N., García-Madruga, J., Gutiérrez, F., & Moreno, S. (1999). How does content affect “unless” conditional reasoning? In *Proceedings of the European Conference on Cognitive Science* (pp. 271–277), Siena, Italy.
- Case, R. (1985). *Intellectual development: Birth to adulthood*. New York: Academic Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Elosúa, M. R., Gutierrez, F., García Madruga, J. A., Luque, J. L., & Garate, M. (1996). Versión española del “Reading Span Test” de Daneman and Carpenter. [Spanish version of Daneman & Carpenter’s “Reading Span Test”]. *Psicothema*, 2, 383–395.
- Engle, R. W., Kane, M. J., & Tuholsky, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 102–134). Cambridge, UK: Cambridge University Press.

- Engle, R. W., & Oransky, N. (1999). Multi-store versus dynamic models of temporary storage in memory. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 515–555). Cambridge, MA: MIT Press.
- Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, 75, 451–468.
- Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. London: Lawrence Erlbaum Associates Ltd.
- Evans, J. St. B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition*, 48, 1–20.
- Evans, J. St. B. T., & Over, D. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.
- Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition*, 77, 1–79.
- Fillenbaum, S. (1986). The use of conditionals in inducements and deterrents. In E. C. Traugott, A. ter Meulen, J.S. Reilly, & C.A. Ferguson (Eds.), *On conditionals* (pp. 179–195). Cambridge, UK: Cambridge University Press.
- Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, 51, 136–158.
- García Madruga, J. (1983). Un modelo general sobre el razonamiento silogístico: Doble procesamiento y fase de comprobación con verificación [A general model on syllogistic reasoning: Dual processing and a verification test-phase]. *Revista de Psicología General y Aplicada*, 3, 439–466.
- García-Madruga, J. A., Gárate, M., Elosúa, R., Luque, J. L., & Gutiérrez, F. (1997). Comprensión lectora y memoria operativa: un estudio evolutivo. [Reading comprehension and working memory: a developmental study]. *Cognitiva*, 9(1), 99–132.
- García-Madruga, J. A., Gutiérrez, F., Carriero, N., Luzón J. M., & Vila, J. O. (2005). Working memory and propositional reasoning: Searching for new working memory tests. In V. Girotto & P. N. Johnson-Laird (Eds.), *The shape of reason. Essays in honour of Paolo Legrenzi*. London: Psychology Press.
- García-Madruga, J. A., Gutiérrez, F., Carriero, N., Moreno, S., & Johnson-Laird, P. N. (2002). Mental models in deductive reasoning. *The Spanish Journal of Psychology*, 5, 125–140.
- García-Madruga, J. A., Moreno, S., Carriero, N., Gutiérrez, F., & Johnson-Laird, P. N. (2001). Are conjunctive inferences easier than disjunctive inferences? A comparison of rules and models. *Quarterly Journal of Experimental Psychology*, 54A, 613–632.
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove, UK: Psychology Press.
- Gathercole, S., & Baddeley, A. D. (1993). *Working memory and language*. Hove, UK: Lawrence Erlbaum Associates Inc.
- Geis, M. L. (1973). If and unless. In B. B. Kachru, R. B. Lees, Y. Malkiel, A. Pietrangeli, & S. Saporta (Eds.), *Issues in linguistics* (pp. 231–253). Urbana Chicago: University of Illinois Press.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Earlbaum Associates Inc.
- Gilhooly, K. J. (1998). Working memory, strategies, and reasoning task. In R. H. Logie, & K. J., Gilhooly (Eds.), *Working memory and thinking* (pp. 23–45). Hove, UK: Psychology Press.
- Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Journal of Memory and Cognition*, 21(1), 115–124.
- Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *European Journal of Cognitive Psychology*, 11(4), 473–498.

- Gutiérrez, F., García-Madruga, J. A., Carriero, N., Luzón, J. M., & Vila, J. O. (2005). Dos pruebas de Amplitud de Memoria Operativa para el Razonamiento [Two tests of Working Memory Span for Reasoning]. *Cognitiva*, 17, 183–207.
- Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. T. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning*, 10, 175–195.
- Holyoak, K., & Cheng, P. (1995). Pragmatic reasoning with a point of view. *Thinking & Reasoning*, 1, 289–400.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, UK: Cambridge University Press.
- Johnson-Laird, P. N. (1985). Deductive reasoning ability. In R. J. Sternberg (Ed.), *Human abilities: An information process approach* (pp. 173–194) New York: W. H. Freeman.
- Johnson-Laird, P. N. (2000). The current state of the mental model theory. In J. A. García-Madruga, N. Carriero, & M. J. González Labra (Eds.), *Mental models in reasoning* (pp. 16–40). Madrid: UNED.
- Johnson-Laird, P. N. (2004). The history of mental models. In K. Manktelow & P. Cheung Chun (Eds.), *Psychology of reasoning*. Hove, UK: Psychology Press.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates Inc.
- Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646–678.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension. *Psychological Review*, 99, 122–149.
- Klauer, K. C., Stegmaier, R., & Meiser, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking and Reasoning*, 3, 9–48.
- Markovits, H., Doyon, C., & Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. *Thinking and Reasoning*, 8, 97–107.
- Meilán, E., García-Madruga, J. A., & Vieiro, P. (2000). Memoria operativa y procesos de razonamiento condicional. [Working memory and conditional reasoning processes]. *Cognitiva*, 12, 135–151.
- Morris, B. J., & Sloutsky, V. (2002). Children's solutions of logical versus empirical problems: What's missing and what develops? *Cognitive Development*, 16, 907–928.
- Norman, D. A., & Shallice, T. (1980). *Attention to action: Willed and automatic control of behavior*. University of California at San Diego, CHIP Report 99.
- Philips, L. H., & Forshaw, M. J. (1998). The role of working memory in age differences in reasoning. In R. H. Logie, & K. J. Gilhooly (Eds.), *Working memory and thinking* (pp. 23–45). Hove, UK: Psychology Press.
- Quine, W. V. (1972). *Methods of logic* (3rd ed.). New York: Holt, Rinehart & Winston.
- Reichenbach, H. (1947). Philosophy: Speculation or science? *Nation*, 164, 19–22.
- Richardson, J., & Ormerod, T. C. (1997). Rephrasing between disjunctives and conditionals: Mental models and the effects of thematic content. *Quarterly Journal of Experimental Psychology*, 50A, 358–385.
- Richardson, J. T. E. (1996). Evolving concepts of working memory. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zacks (Eds.), *Working memory and human cognition* (pp. 3–30). New York: Oxford University Press.
- Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human reasoning*. Cambridge, MA: MIT Press.
- Schaeken, W., García-Madruga, J. A., & D'Ydewalle, G. (1997). *Unless reasoning*. Unpublished manuscript. Laboratory of Experimental Psychology, Leuven University, Belgium.

- Shallice, T., & Burgess, P. (1993). Supervisory control of action and thought selection. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention, selection, awareness, and control: A tribute to Donald Broadbent* (pp. 171–187). Oxford, UK: Oxford University Press.
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Toms, M., Warris, N., & Ward, D. (1993). Working memory and conditional reasoning. *Quarterly Journal of Experimental Psychology*, 46A, 679–699.
- Whitney, P., Arnett, P. A., Driver, A., & Budd, D. (2001). Measuring central executive functioning: What's in a reading span? *Brain and Cognition*, 45, 1–14.
- Whitney, P., Ritchie, B., & Clark, M. (1991). Working memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes*, 14, 133–145.

**Copyright of Thinking & Reasoning is the property of Psychology Press (UK) and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.**